

Statistics

Samples, estimation, inference, and testing

Last modified: 2026-10-18

i Source and scope

This appendix outlines the statistical foundations and terminology assumed throughout these notes. It points to the definitions, worked calculations, and examples in the lab's shared prerequisite site: *Statistics for Data Science*^a.

Machine learning and statistics share much of their mathematical foundations, but often differ in their primary goals: statistics emphasizes inference, parameter estimation, and hypothesis testing, whereas machine learning emphasizes generalization and predictive accuracy on unseen data [1], [2].

^a<https://morrison-lab.github.io/sds/>

Statistical theory explains what conclusions a finite sample of data supports, how to estimate underlying model parameters, and how to measure uncertainty. This appendix outlines the core concepts and links directly to their full definitions in *Statistics for Data Science*¹.

1 Populations, samples, and empirical distributions

*Statistics for Data Science*² covers the distinction between populations and finite samples:

- Populations³ and samples⁴: A population distribution describes the entire data-generating process, while a sample consists of N observed data points drawn from that population.
- Independent and identically distributed⁵ (i.i.d.) observations: Sampling data independently from a shared underlying distribution $p(\tilde{x}, y)$.
- Empirical distributions⁶: Placing probability mass $1/N$ on each observed training instance, which forms the basis for empirical risk minimization [3].
- Standardization⁷ and min-max scaling⁸: Centering features to mean zero and scaling to unit variance (z-scores), or scaling to a fixed interval $[0, 1]$.

i Watch

Starmer explains data scaling and normalization^a [4].

^ahttps://www.youtube.com/watch?v=s_aT_lQ_e3w

¹<https://morrison-lab.github.io/sds/>

²<https://morrison-lab.github.io/sds/>

³<https://morrison-lab.github.io/sds/exploratory-descriptive.html#def-population>

⁴<https://morrison-lab.github.io/sds/exploratory-descriptive.html#def-sample>

⁵<https://morrison-lab.github.io/pds/independence.html#def-iid>

⁶<https://morrison-lab.github.io/sds/nonparametric-models.html#def-empirical-distribution>

⁷<https://morrison-lab.github.io/sds/exploratory-descriptive.html#def-standardization>

⁸<https://morrison-lab.github.io/sds/exploratory-descriptive.html#def-min-max-scaling>

i Read

GeeksforGeeks explains:

- feature scaling^a [5]
- exploratory data analysis^b [6]

^a<https://www.geeksforgeeks.org/data-science/feature-scaling-in-machine-learning/>

^b<https://www.geeksforgeeks.org/data-science/what-is-exploratory-data-analysis/>

2 Point estimation and estimator properties

Statistics for Data Science covers parameter estimation and estimator performance:

- Point estimators⁹: A rule or statistic $\hat{\theta} = g(X_1, \dots, X_N)$ used to estimate an unknown population parameter θ [7].
- Estimator bias¹⁰: $\text{Bias}(\hat{\theta}) \triangleq \mathbb{E}[\hat{\theta}] - \theta$. An estimator is unbiased when $\text{Bias}(\hat{\theta}) = 0$ [8].
- Estimator variance and standard error¹¹: $\text{Var}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$, with standard error $\text{SE}(\hat{\theta}) \triangleq \sqrt{\text{Var}(\hat{\theta})}$.
- Mean squared error¹² (MSE) and the bias-variance decomposition of an estimator¹³: $\text{pMSE}(\hat{\theta}) \triangleq \mathbb{E}[(\hat{\theta} - \theta)^2] = \text{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta})$.
- Maximum likelihood estimation¹⁴ (MLE): Estimating parameters by maximizing the likelihood function $L(\theta; \tilde{x}) = \prod_{i=1}^N p(x_i; \theta)$, or equivalently the log-likelihood $\ell(\theta; \tilde{x}) = \sum_{i=1}^N \ln p(x_i; \theta)$.

i Watch

Starmer explains:

- Maximum Likelihood Estimation^a [9]
- the bias-variance tradeoff^b [10]

^a<https://www.youtube.com/watch?v=XepXtl9YKwc>

^b<https://www.youtube.com/watch?v=EuBBz3bI-aA>

i Read

GeeksforGeeks explains:

- maximum likelihood estimation^a [11]
- the bias-variance tradeoff^b [12]

^a<https://www.geeksforgeeks.org/data-science/maximum-likelihood-estimation/>

^b<https://www.geeksforgeeks.org/maths/bias-variance-trade-off/>

i Watch

5MinuteAI explains:

- maximum likelihood estimation^a, with examples^b
- properties of estimators^c (the first of three videos)

Both come from its Probability & Statistics^d playlist [13].

^a<https://www.youtube.com/watch?v=Pa-GWE5JuH4>

⁹<https://morrison-lab.github.io/sds/estimation.html#def-estimator>

¹⁰<https://morrison-lab.github.io/sds/estimation.html#def-bias>

¹¹<https://morrison-lab.github.io/sds/estimation.html#def-SE>

¹²<https://morrison-lab.github.io/sds/estimation.html#def-mse>

¹³<https://morrison-lab.github.io/sds/estimation.html#thm-mse-bias-variance>

¹⁴<https://morrison-lab.github.io/sds/intro-MLEs.html#def-mle>

^b<https://www.youtube.com/watch?v=w3a08digzfw>

^c<https://www.youtube.com/watch?v=sUFpzYNDn3Q>

^d<https://www.youtube.com/playlist?list=PLb45KifGiuHesi4PALNZSYZFhViVGQJK>

i Read

The University of Washington's CSE 446 slides cover maximum likelihood estimation^a [14], Lecture 2.

^a<https://courses.cs.washington.edu/courses/cse446/26sp/schedule/lectures/02-maximum%20likelihood%20estimation.pdf>

3 Hypothesis testing and confidence intervals

Statistics for Data Science covers formal hypothesis testing and uncertainty quantification:

- Null hypothesis¹⁵ (H_0 , the default baseline or null effect) versus alternative hypothesis¹⁶ (H_1 , the alternative claim).
- Test statistics and p -values¹⁷: A test statistic summarizes sample evidence, and the p -value is the probability under H_0 of observing a test statistic at least as extreme as the one calculated.
- Significance levels¹⁸ and decision errors: A test with significance level α rejects H_0 when the p -value is at most α . A Type I error (false positive) occurs when rejecting a true H_0 ; a Type II error (false negative) occurs when failing to reject a false H_0 , with statistical power defined as $1 - \beta$.
- Confidence intervals¹⁹: Constructing an interval $[L, U]$ that covers the true parameter θ with probability $1 - \alpha$ under repeated sampling [15].

i Watch

5MinuteAI explains confidence intervals^a and hypothesis testing^b [13].

^a<https://www.youtube.com/watch?v=XdaAhYUqWsg>

^bhttps://www.youtube.com/watch?v=Plya_ZXkOvs

i Read

GeeksforGeeks explains:

- hypothesis testing^a [16]
- confidence intervals^b [17]

^a<https://www.geeksforgeeks.org/data-science/understanding-hypothesis-testing/>

^b<https://www.geeksforgeeks.org/maths/confidence-interval/>

4 Statistical inference versus machine learning prediction

Machine learning and statistics share data and mathematical tools, but often approach modeling from different perspectives:

- Inference²⁰: Focuses on understanding the data-generating mechanism, estimating interpretable parameters, quantifying uncertainty via standard errors²¹ and confidence intervals²², and testing specific

¹⁵<https://morrison-lab.github.io/sds/inference.html#def-null-hypothesis>

¹⁶<https://morrison-lab.github.io/sds/inference.html#def-alternative-hypothesis>

¹⁷<https://morrison-lab.github.io/sds/inference.html#def-p-value>

¹⁸<https://morrison-lab.github.io/sds/inference.html#def-significance-level>

¹⁹<https://morrison-lab.github.io/sds/inference.html#def-confidence-interval>

²⁰<https://morrison-lab.github.io/sds/inference.html#def-inference>

²¹<https://morrison-lab.github.io/sds/estimation.html#def-SE>

²²<https://morrison-lab.github.io/sds/inference.html#def-confidence-interval>

scientific hypotheses²³.

- Prediction²⁴: Focuses on constructing models that generalize accurately to new, unseen observations, often accepting complex, regularized, or non-parametric models where individual parameters may not have a simple mechanistic interpretation.
- Course Overview discusses the relationship between machine learning and statistics in detail, and provides a translation table between the vocabularies of both fields.
- **Historical and cultural roots**: The divergence in terminology and modeling philosophy is surveyed in classic papers by Breiman [18] on the algorithmic and data modeling cultures, Ripley [19], Table 1.1 on parallel terms across neural networks and statistics, Varian [20] comparing machine learning and econometrics, and Efron and Hastie [21] on computer age statistical inference.

5 Further reading

- James et al. [1], Chapter 2
- Hastie et al. [2], Chapters 1 and 2
- Starmer [22], on statistical testing and estimation

References

- [1] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning with applications in R*, 2nd ed. Springer, 2021. Available: <https://www.statlearning.com/>
- [2] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: Data mining, inference, and prediction*, 2nd ed. Springer, 2009. Available: <https://hastie.su.domains/ElemStatLearn/>
- [3] Wikipedia contributors, “Empirical distribution function — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Empirical_distribution_function
- [4] J. Starmer, “One-hot, label, target and k-fold target encoding, clearly explained.” StatQuest, video on YouTube, 2023. Available: <https://www.youtube.com/watch?v=589nCGeWG1w>
- [5] GeeksforGeeks, “Feature engineering - scaling, normalization and standardization.” GeeksforGeeks, tutorial article, 2026. Available: <https://www.geeksforgeeks.org/machine-learning/feature-engineering-scaling-normalization-and-standardization/>
- [6] GeeksforGeeks, “Exploratory data analysis.” GeeksforGeeks, tutorial article, 2026. Available: <https://www.geeksforgeeks.org/data-analysis/what-is-exploratory-data-analysis/>
- [7] Wikipedia contributors, “Estimator — Wikipedia, the free encyclopedia.” 2026. Available: <https://en.wikipedia.org/wiki/Estimator>
- [8] Wikipedia contributors, “Bias of an estimator — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Bias_of_an_estimator
- [9] J. Starmer, “Maximum likelihood, clearly explained.” StatQuest, video on YouTube, 2017. Available: <https://www.youtube.com/watch?v=XepXtl9YKwc>
- [10] J. Starmer, “Machine learning fundamentals: Bias and variance.” StatQuest, video on YouTube, 2018. Available: <https://www.youtube.com/watch?v=EuBBz3bI-aA>
- [11] GeeksforGeeks, “Probability density estimation & maximum likelihood estimation.” GeeksforGeeks, tutorial article, 2025. Available: <https://www.geeksforgeeks.org/machine-learning/probability-density-estimation-maximum-likelihood-estimation/>
- [12] GeeksforGeeks, “Bias-variance trade off - machine learning.” GeeksforGeeks, tutorial article, 2026. Available: <https://www.geeksforgeeks.org/machine-learning/ml-bias-variance-trade-off/>
- [13] 5MinuteAI, “Probability & statistics.” Video playlist on YouTube. Accessed: Oct. 12, 2026. [Online]. Available: <https://www.youtube.com/playlist?list=PLB45KifGiuHesi4PALNZSYZfHViVGQJK>
- [14] N. Jaques, “CSE 446/546: Machine learning, lecture slides.” University of Washington, 2026. Accessed: Oct. 12, 2026. [Online]. Available: <https://courses.cs.washington.edu/courses/cse446/26sp/schedule/lectures/>

²³<https://morrison-lab.github.io/sds/inference.html#def-null-hypothesis>

²⁴<https://morrison-lab.github.io/sds/prediction.html#def-prediction>

- [15] Wikipedia contributors, “Confidence interval — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Confidence_interval
- [16] GeeksforGeeks, “Hypothesis testing.” GeeksforGeeks, tutorial article, 2025. Available: <https://www.geeksforgeeks.org/data-science/understanding-hypothesis-testing/>
- [17] GeeksforGeeks, “Confidence interval.” GeeksforGeeks, tutorial article, 2025. Available: <https://www.geeksforgeeks.org/maths/confidence-interval/>
- [18] L. Breiman, “Statistical modeling: The two cultures,” *Statistical Science*, vol. 16, no. 3, 2001, doi: 10.1214/ss/1009213726²⁵.
- [19] B. D. Ripley, *Pattern recognition and neural networks*. Cambridge: Cambridge University Press, 1996. doi: 10.1017/CBO9780511812651²⁶.
- [20] H. R. Varian, “Big data: New tricks for econometrics,” *Journal of Economic Perspectives*, vol. 28, no. 2, pp. 3–28, 2014, doi: 10.1257/jep.28.2.3²⁷.
- [21] B. Efron and T. Hastie, *Computer age statistical inference: Algorithms, evidence, and data science*. Cambridge: Cambridge University Press, 2016. doi: 10.1017/CBO9781316576533²⁸.
- [22] J. Starmer, *The StatQuest illustrated guide to statistics: With hands-on examples in Python and R*. 2026. Available: <https://statquest.org/statquest-store/>

²⁵<https://doi.org/10.1214/ss/1009213726>

²⁶<https://doi.org/10.1017/CBO9780511812651>

²⁷<https://doi.org/10.1257/jep.28.2.3>

²⁸<https://doi.org/10.1017/CBO9781316576533>